AWARD NUMBER: W81XWH-11-1-0713

TITLE: Identification of Novel, Inherited Genetic Markers for Aggressive PCa in European and African Americans Using Whole Genome Sequencing

PRINCIPAL INVESTIGATOR: Jielin Sun, Ph.D.

CONTRACTING ORGANIZATION: Wake Forest University Health Sciences
Winston Salem, NC 27157

REPORT DATE: September 2012

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 September 2012	Annual	22 Aug 2011 – 21 Aug 2012

4. TITLE AND SUBTITLE

Identification of Novel, Inherited Genetic Markers for Aggressive PCa in European and African Americans Using Whole Genome Sequencing

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0713

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jielin Sun, Ph.D.

Email: jisun@wakehealth.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wake Forest University Health Services
Winston Salem, NC 27157

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Prostate cancer (PCa) is the most common cancer and the second leading cause of cancer death among men in the United States. While most prostate cancer (PCa) patients have an indolent form of the disease that may not even require treatment, about 10-15% of PCa patients have an aggressive form that may progress to metastases and death thus requiring intensive treatment. Several clinical variables such as PSA levels, Gleason grade, and TNM stage are good predictors for disease with poor clinical outcomes; however, their predictive performance needs to be improved. Our inability to reliably distinguish between these two forms of PCa, early on in the course of the disease has resulted in the over-treatment of many and under treatment of some. The identification of additional markers, including genetic variants will improve our ability to distinguish aggressive from indolent forms of PCa and to better understand the racial disparity of PCa that exists between EAs and AAs. In this DOD proposal, we hypothesized that multiple rare sequence variants in the genome may increase aggressive PCa risk. Through a genome-wide search of rare variants based on an existing population from Johns Hopkins Hospital (JHH) of 400 aggressive PCa and 400 indolent PCa using Illumina Human Exome BeadChip, we identified several rare variants that are significantly associated with aggressive PCa development in EA or AA populations. The implicated rare variants will be followed in additional populations.

15. SUBJECT TERMS
Prostate cancer, indolent, lethal (aggressive), sequence variants

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	12	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

While most prostate cancer (PCa) patients have an indolent form of the disease that may not even require treatment, about 10-15% of PCa patients have an aggressive form that may progress to metastases and death thus requiring intensive treatment. Several clinical variables such as PSA levels, Gleason grade, and TNM stage are good predictors for disease with poor clinical outcomes; however, their predictive performance needs to be improved. Our inability to reliably distinguish between these two forms of PCa, early on in the course of the disease has resulted in the over-treatment of many and under treatment of some. Another dilemma is a large difference in PCa risk, especially aggressive PCa, between races. African Americans (AAs) have the world's highest incidence of PCa and are twice as likely, as compared with Caucasians to die of the disease. Inherited markers of aggressive PCa could be used for screening and diagnosis of aggressive PCa at an early stage while reducing over-diagnosis and treatment for others. The overall hypothesis is that inherited sequence variants in the genome are associated with a lethal (aggressive) form of PCa but not indolent PCa, and the difference in these variants between races may contribute to higher incidence of and mortality from aggressive PCa in AA.

In this DOD proposal, we propose to identify 1) To discover novel inherited genetic variants in the genome that may be associated with aggressive but not indolent PCa using a whole genome sequencing (WGS) approach in hereditary PCa families (HPC); **2)** To confirm the novel genetic variants using mass spectrometry directed sequencing; and **3)** To perform association tests of implicated genetic variants among 1,500 most aggressive PCa and 1,500 least aggressive (i.e. indolent) PCa.

BODY

Approved Statement of Work:

Statement of Work
Aim 1. To discover novel inherited genetic variants in the genome that may be associated with aggressive but not indolent PCa using a WGS approach.

Step by Step method and expected results

1. **Months 1-6:** Preparation of the study, including regulatory review, IRB approval and other logistical issues
2. **Months 7-12**: Perform a WGS analysis using NGS among 20 men from HPC families.
3. **Months 13-18:** Apply several filter criteria to identify a subset of mutations that most likely associate with aggressive PCa only by: 1) mutations that segregate with aggressive PCa but not indolent PCa or unaffected men in families, and 2) mutations that in functional regions of the genome based on bioinformatics analysis..

Outcome and deliverables

We expect to identify a certain number (~1,000) of novel variants that most likely associate with aggressive but not indolent PCa.

Aim 2. To confirm the genetic variants implicated in Aim 1 using Sequenom

Step by Step method and expected results

1. **Months 19-22:** Genotyping ~1,000 SNPs among 20 samples using Sequenom
2. **Months 22-24**: Confirmation analysis of the ~1,000 SNPs

Outcome and deliverable
We expect that a subset of these 1,000 SNPs will be confirmed using Sequenom platform.

Aim 3. To perform association tests of selected genetic variants among 1,500 more aggressive PCa and 1,500 most indolent PCa.

Step by Step method and expected results

1. **Months 25-26:** Genotyping ~100 SNPs in 1,500 more aggressive PCa and 1,500 most indolent PCa patients
2. **Months 27-28:** Association test of these SNPs with aggressiveness of PCa using a logistic regression model
3. **Months 29-36:** Final analysis and preparation of papers

Outcome and deliverable

We expect that several novel SNPs that are identified through WGS will be associated with aggressiveness of PCa. We will prepare and submit papers reporting the major results from the study.

Summary report

By Sep 2012, we were in the 12th month of this funded project. During the last year, we have completed the following 1) IRB and other logistical issues, 2) performed genotyping of exome-array among 400 aggressive PCa and 400 indolent PCa in European American (EA) and AA (African American) samples, 3) performed single rare variant analysis, bioinformatics analysis, as well as gene-based analysis (SKAT) to identify rare variants that have strong effects on aggressive PCa risk.

Detailed report

Study design modification. In our initial report, we proposed to conduct whole-genome sequencing for 20 patients in the Johns Hopkins Hospital (JHH) population, including 10 EAs and 10 AAs. However, after reviewing the latest published literatures between the date we submitted the original proposal and the actual project start date, we felt the proposed method was no longer the most cost-effective approach to identify novel genetic variants that confer risk to aggressive PCa.

Successful example of rare mutations. Recently, rare mutations (MAF<5%) have been shown to confer large effects to PCa and aggressive PCa. One of the most significant findings in 2012 was the identification of a rare variant on the HOXB13 gene (Ewing 2012, Arbari 2012, Kalsson 2012), with a large effect (OR of 2.0-4.0) to PCa. In addition, rare mutations in BRCA2 and aggressive PCa have also been recently reported. A population-based study conducted using Ashkenazi Jewish population discovered that carriers of the BRCA2 6174delT, had more advanced tumor stage, higher tumor grade and shorter median PCa specific survival time, compared to non-carriers (Gallagher 2010). Another study conducted in Australia that evaluated 26 unique mutations in BRCA2 found similar conclusion (Thorne 2011). Another study recently reported significant differences in histologic grade (Gleason score ≥8, 50% versus 21%), tumor stage (T≥3, 62% versus 18%), nodal diseases (35% versus 11%) and metastasis (21% versus 9%) for BRCA2 mutation carriers and non-carriers, respectively (Castro 2011). These findings provide evidence for the impact of rare mutations on PCa aggressiveness, and the effect is much larger than that of common SNPs contributing to aggressive PCa. However, our knowledge about rare variants and aggressive PCa remain limited, and systematic studies of rare variants, including genome-wide evaluations for aggressive PCa have not been conducted yet. ***Therefore, we propose to modify our study design to identify rare mutations in the genome that are associated with aggressive PCa. However, we won't be able to study such rare variants in our initial design because of the higher cost of whole-genome sequencing and low power to detect such rare mutations based on a small sample size of 20. For example, with 20 samples proposed using whole-genome sequencing, the rare variants with a MAF less than 5% would not be observed. Therefore, we would like to study such rare variants using the newest Illumina Exome BeadChip platform.***

Justification of Using Exome SNP Array to perform our study. The Illumina Human Exome BeadChip became available in early 2012 and represented the newest gene chip that delivers unparalleled coverage of

putative functional exonic variants. The relatively low cost makes it possible to study larger sample sizes. The Exome Beachip is comprised of >240,000 markers, including >200,000 nonsynonymous SNPs, nonsense mutations, SNPs in splice sites and promoter regions, as well as thousands of GWAS tag markers. Nearly 90% of the SNPs on the exome arrays are rare, with a MAF<5%. In addition, the markers on Illumina Human Exome BeadChips are selected from over 12,000 individual exome and whole-genome sequences, representing diverse populations, including those of European and African descent. ***Therefore, it is more efficient and economical to use exome arrays to identify rare variants associated with aggressive PCa, compared with whole genome sequencing.*** We will be able to genotype a total of 600 Aggressive PCa and 600 indolent PCa, including 300 Aggressive PCa and 300 indolent PCa of EAs, as well as 300 Aggressive PCa and 300 indolent PCa utilizing this new technology. We have >80% power to detect an OR of 2.0 (3.6) for variants with a MAF of 0.05 (0.01), at an alpha level of 1E-05 (2-sided).

Study population. The study samples were selected from a hospital-based study population collected at JHH. De-identified DNA samples are available at present. By Dec. 2011, DNA samples from 9,622 patients have been successfully isolated from normal seminal vesicle tissues, including 8,796 EA and 826 AA (Table 1). A unique advantage of this cohort is that all tumors have been uniformly graded and staged based on radical surgery specimens. In this aim, we randomly selected 300 aggressive patients (defined as Gleason score ≥ 4+3, or stage ≥ T3b, or PSA ≥ 20 ng/mL) and 300 indolent patients of EAs, as well as 300 aggressive patients and 300 indolent patients of AAs.

Bioinformatics analysis and statistical analysis

a) Variant effect prediction: All coding nonsynonymous variants were assessed for potential effect by Polymorphism Phenotyping version 2 (PolyPhen2), which is a tool for predicting the possible impact of an amino acid substitution on the structure and function of a human protein. For a given variant, PolyPhen2 calculates a Naïve Bayes posterior probability that the mutation is damaging and then appraised qualitatively as benign, possibly damaging, or probably damaging (Adzhubei 2010).

b) Single variant analysis: Logistic regression was performed to test association between each rare variant and aggressive PCa, adjusting for age. If the expected number of mutations is smaller than 5, a Fisher's exact test was used.

c) Gene-based analysis: We used SKAT, to conduct gene-based analysis of rare variants for aggressive PCa. SKAT is a supervised and flexible regression method to test for association between rare variants in a gene or genetic region and a continuous or dichotomous trait. Compared to other methods of estimating the joint effect of a subset of SNPs, SKAT is able to deal with variants that have different direction and magnitude of effects, and allows for covariate adjustment (Wu 2011). In addition, SKAT can also avoid arbitrary selection of threshold in burden test. Moreover, SKAT is computationally efficient, compared to a permutation test, making it convenient to analyze the large dataset in our study.

Results

EA population

The top significant SNPs that were significantly associated with aggressive PCa in EAs are listed in Table 1 (200 aggressive cases vs 200 nonaggressive cases). A total of 35 SNPs were included in Table 1 with P-value < 1E-03. The top significant SNP, rs114000606, was located on the UBIAD1 gene on chromosome 1, with a MAF of 0.028 in aggressive PCa and 0 in indolent PCa.

We then performed gene-based analysis using the SKAT approach. The top 45 genes with P-value < 5-05 are presented in Table 2. The UBIAD1 gene was also identified as the most significant gene associated with aggressive PCa, with a P-value of 3.3E-06.

Table 1. Top signficant variants associated with aggressive PCa in EAs from JHH population

SNP	CHR	BP	A1	A2	Maf_case	Maf_ctrl	P	OR	Category	Gene Name
rs114000606	1	11,333,812	A	G	0.02817	0	7.08E-07		missense	*UBIAD1*
rs10057851	5	64,565,261	G	A	0.4233	0.528	3.78E-05	0.656	Intron	*ADAMTS6*
rs4903104	14	73,735,366	A	G	0.2161	0.1384	5.23E-05	1.717	missense	*PAPLN*
rs36101975	18	28,956,904	A	G	0.09773	0.1671	6.72E-05	0.5401	silent	*DSG4*
rs35833603	3	41,973,460	C	G	0.002817	0.02784	1.08E-04	0.09863	missense	*ULK4*
rs6934690	6	54,054,686	T	A	0.04085	0.08933	1.36E-04	0.4341	missense	*MLIP*
rs7564372	2	85,590,286	A	G	0.008451	0.03828	1.54E-04	0.2141	missense	*ELMOD3*
rs41271546	6	29,323,245	G	C	0.02254	0.00232	1.78E-04	9.914	missense	*OR5V1*
rs3752095	18	28,934,681	T	A	0.09859	0.1624	2.14E-04	0.5641	missense	*DSG1*
rs4679904	3	160,340,896	A	G	0.3254	0.2419	2.44E-04	1.512	Intergenic	*ARL14*
rs2275769	6	54,095,524	A	G	0.04085	0.08701	2.48E-04	0.4469	missense	*MLIP*
rs7557290	2	77,512,571	A	G	0.01831	0.05349	2.64E-04	0.33	Intron	*LRRTM4*
rs61898615	11	103,019,260	A	G	0.02676	0.00464	2.78E-04	5.898	missense	*DYNC2H1*
rs7966162	12	77,154,974	G	A	0.3545	0.4443	3.08E-04	0.6869	Intergenic	*ZDHHC17*
rs2069541	14	23,901,012	G	A	0.02113	0.00232	3.34E-04	9.281	silent	*MYH7*
rs2069541	14	23,901,012	G	A	0.02113	0.00232	3.34E-04	9.281	silent	*MYH7*
rs2232548	12	9,985,915	A	C	0.1017	0.05452	4.42E-04	1.963	missense	*KLRF1*
rs56138314	5	173,426,709	A	C	0.04661	0.01624	4.42E-04	2.961	missense	*C5orf47*
rs34613961	19	18,701,700	A	G	0.01408	0	4.73E-04		missense	*C19orf60*
rs34795598	18	29,848,028	G	A	0.07746	0.03712	4.99E-04	2.178	missense	*FAM59A*
bs2_192701301	2	192,701,301	A	C	0.02535	0.00464	5.03E-04	5.579	missense	*SDPR*
rs10771604	12	30,098,770	A	G	0.3817	0.4687	5.25E-04	0.6998	Intergenic	*TMTC1*
rs395136	19	33,106,742	G	A	0.1831	0.1206	5.37E-04	1.634	Intron	*ANKRD27*
rs12023499	1	155,031,376	A	G	0.2458	0.1752	5.93E-04	1.534	Intron	*LOC100505666*
rs13031237	2	61,136,129	A	C	0.3778	0.2958	6.11E-04	1.446	Intron	*REL*
rs8192646	6	132,938,842	A	G	0.03944	0.01276	7.15E-04	3.176	nonsense	*TAAR2*
rs4909945	11	10,673,739	A	G	0.2778	0.3581	7.22E-04	0.6893	missense	*MRVI1*
rs78403475	9	139,235,606	C	G	0.1116	0.06395	7.87E-04	1.838	missense	*GPSM1*
rs8096726	18	47,837,090	A	G	0.4547	0.3712	8.26E-04	1.412	Intergenic	*CXXC1*
rs12948945	17	75,038,439	A	C	0.4479	0.5326	8.38E-04	0.712	Intergenic	*SCARNA16*
rs20541	5	131,995,964	A	G	0.2238	0.1578	8.61E-04	1.539	missense	*IL13*
rs753414	15	45,474,371	C	A	0.3141	0.3951	8.68E-04	0.701	Intron	*SHF*
bs6_139609687	6	139,609,687	A	G	0.01268	0	9.16E-04		missense	*TXLNB*
rs3808795	9	17,273,731	G	A	0.2259	0.3002	9.47E-04	0.68	missense	*CNTLN*
rs432869	22	21,408,430	A	G	0.5	0.4165	9.50E-04	1.401	Intron	*LOC400891*

Table 2. Top signficant genes associated with aggressive PCa using SKAT approach in EAs from JHH population

Gene	P-value
UBIAD1	3.3E-06
MLIP	2.7E-04
LRRTM4	3.1E-04
ELMOD3	3.2E-04
MYH7	3.4E-04
C5orf47	5.3E-04
LOC400891	5.9E-04
SCARNA16	6.8E-04
GML	7.5E-04
LOC100505666	8.0E-04
FAM193A	9.2E-04
COX7B2	1.1E-03
POLB	1.4E-03
ZNF223	1.4E-03
MYH6	1.5E-03
ATP2A3	1.7E-03
TAAR2	1.7E-03
CLEC3B	1.8E-03
SHH	1.9E-03
HOXA7	2.1E-03
LOC100507472	2.1E-03
TMEM233	2.1E-03
ZDHHC21	2.1E-03
ZPBP2	2.3E-03
SEC23B	2.3E-03
SLC16A14	2.4E-03
TENC1	2.5E-03
PINK1	2.6E-03
LGR5	2.8E-03
CNTD2	2.8E-03
DYNC2H1	2.9E-03
KLK5	3.1E-03
TRMT12	3.1E-03
PIN4	3.4E-03
PCDHGA6	3.7E-03
DSG3	3.9E-03
ACTRT2	4.1E-03
LOC730101	4.3E-03
MIR30B	4.4E-03
ADA	4.6E-03
CCDC108	4.6E-03
OR52E6	4.6E-03

SLC38A5	4.9E-03
MOCS3	4.9E-03
ZNF544	4.9E-03

AA population

The top significant SNPs that were significantly associated with aggressive PCa in AAs (200 aggressive cases vs 200 nonaggressive cases) are listed in Table 3. A total of 35 SNPs with P-value < 1E-03 are presented in Table 3. The top significant SNP rs61227179 was located on the ZNF12 gene on chromosome 17, with a MAF of 0.08 in aggressive PCa and 0.026 in indolent PCa.

We then performed gene-based analysis using the SKAT approach. The top 28 genes with P-value < 5-05 are presented in Table 4. The ASB9 gene was identified as the most significant gene and is associated with aggressive PCa, with a P-value of 8.6E-05.

Table 3. Top signficant variants associated with aggressive PCa in AAs from JHH population

SNP	CHR	BP	A1	A2	Maf_case	Maf_ctrl	P	OR	Category	GeneName
rs61227179	7	6,732,315.00	C	A	0.08	0.02648	5.10E-05	3.20	missense	*ZNF12*
rs2291122	23	15,265,457.00	A	G	0.5682	0.3969	8.83E-05	2.00	Intron	*ASB9*
rs2228262	15	39,882,178.00	G	A	0.1556	0.08075	1.10E-04	2.10	missense	*THBS1*
rs77763884	19	4,508,905.00	A	G	0.002222	0.03583	1.93E-04	0.06	missense	*PLIN4*
rs7559772	2	159,166,069.00	A	G	0.01778	0.06542	2.09E-04	0.26	missense	*CCDC148*
rs2294619	16	1,814,440.00	G	A	0.1339	0.2227	2.11E-04	0.54	missense	*MAPK8IP3*
rs4975709	5	1,877,280.00	C	A	0.1956	0.2944	2.20E-04	0.58	Intergenic	*IRX4*
rs78240650	12	91,347,643.00	A	G	0.01111	0.0528	2.53E-04	0.20	missense	*C12orf12*
rs62137612	2	53,994,929.00	A	G	0.03111	0.08567	2.65E-04	0.34	utr5	*CHAC2*
rs115537722	19	4,511,181.00	A	G	0.002222	0.03427	2.84E-04	0.06	missense	*PLIN4*
rs16979912	19	14,910,321.00	G	A	0.1406	0.2281	3.11E-04	0.55	missense	*OR7C1*
rs3803414	15	66,206,204.00	A	G	0.1356	0.07009	3.23E-04	2.08	missense	*MEGF11*
rs6426219	1	247,259,684.00	A	G	0.4085	0.5186	3.36E-04	0.64	Intergenic	*ZNF669*
rs17160911	7	139,138,950.00	G	C	0.1704	0.09783	4.12E-04	1.89	missense	*KLRG2*
rs61750791	13	32,776,616.00	A	T	0.01778	0.06211	4.27E-04	0.27	missense	*FRY*
rs2501340	1	159,824,967.00	G	C	0.1763	0.1028	4.36E-04	1.87	missense	*C1orf204*
rs3806366	1	163,115,321.00	G	A	0.1333	0.07009	4.85E-04	2.04	utr3	*RGS5*
rs73004304	19	14,910,210.00	C	G	0.14	0.2234	5.32E-04	0.57	missense	*OR7C1*
rs6664618	1	66,714,584.00	A	C	0.5045	0.3984	5.36E-04	1.54	Intron	*PDE4B*
rs12366671	12	4,736,569.00	G	A	0.1178	0.1963	5.64E-04	0.55	missense	*AKAP3*
rs1064005	11	33,065,394.00	A	G	0.3067	0.4081	6.16E-04	0.64	silent	*TCP11L1*
rs17230134	19	14,910,654.00	G	A	0.1406	0.2233	6.23E-04	0.57	missense	*OR7C1*
rs118097475	15	89,173,398.00	A	G	0.02667	0.003115	6.61E-04	8.77	missense	*AEN*
rs7752978	6	115,191,061.00	G	A	0.3386	0.4408	7.06E-04	0.65	Intergenic	*HS3ST5*
rs34638481	13	31,891,743.00	A	G	0.05111	0.01553	7.09E-04	3.42	missense	*B3GALTL*
rs6008842	22	46,860,063.00	A	G	0.02232	0.001558	7.39E-04	14.63	missense	*CELSR1*
rs76190154	7	149,493,782.00	A	G	0.03333	0.006231	7.48E-04	5.50	Coding	*SSPO*
rs75995642	2	141,773,450.00	G	A	0	0.02484	7.56E-04	0.00	missense	*LRP1B*
rs41281027	9	104,130,469.00	G	C	0.18	0.109	8.26E-04	1.79	missense	*BAAT*
rs61732336	1	247,752,109.00	G	A	0.1111	0.0559	8.41E-04	2.11	missense	*OR2G2*
rs4317244	4	186,320,906.00	G	C	0.08	0.03416	8.70E-04	2.46	missense	*ANKRD37*

rs10464105	5	180,166,461.00	G	C	0.006667	0.03882	9.20E-04	0.17	missense	*OR2Y1*
rs1036533	2	201,397,724.00	A	G	0.1	0.04829	9.48E-04	2.19	missense	*SGOL2*
rs2271761	2	180,311,444.00	G	A	0.03556	0.007764	9.79E-04	4.71	missense	*ZNF385B*
rs12800642	11	55,339,676.00	A	C	0.2995	0.2118	9.95E-04	1.59	missense	*OR4C16*

Table 4. Top signficant genes associated with aggressive PCa using SKAT approach in AAs from JHH population

Gene	P-value
ASB9	8.6E-05
C12orf12	2.8E-04
PLIN4	4.1E-04
ANKRD37	6.2E-04
OR2G2	1.2E-03
ZNF385B	1.6E-03
HTR7	1.9E-03
TTI1	1.9E-03
AEN	1.9E-03
SOX6	2.0E-03
MIR128-2	2.2E-03
DAND5	2.6E-03
C7orf52	2.8E-03
ATP5SL	2.8E-03
VCX3A	3.0E-03
PTN	3.2E-03
TWF1	3.3E-03
SGSH	3.3E-03
LOC253573	3.4E-03
VBP1	3.5E-03
AZI1	3.6E-03
DAZL	3.7E-03
LOC100506207	3.9E-03
BCL2L14	4.0E-03
CHCHD7	4.2E-03
METTL12	4.3E-03
LRIG1	4.5E-03
OR2Y1	4.7E-03

Discussion

To our knowledge, our study represents one of the first comprehensive studies to identify rare variants that are associated with aggressive PCa in both EAs and AAs. Our data generated from the first year showed potentially important rare variants that are associated with aggressive PCa.

The top rare SNP implicated is a nonsynonymous SNP located on the UBIAD1 gene. The UBIAD1 (TERE1) gene was previously showed to affect growth regulation in prostate carcinoma (McGarvey et al).

The TERE1 gene maps to chromosome 1p36.11-1p36.33, a chromosome locus that has been identified by loss of heterozygosity studies as a site of a putative tumor suppressor gene or genes for multiple tumor types including prostate carcinoma. A significant (61%) decrease in the TERE1 transcript in prostate carcinoma (CaP) and a distinct loss of the TERE1 protein in metastatic prostate cancer was observed in a previous study. Additionally, microarray analysis also showed various growth regulatory genes that are down-regulated or up-regulated in TERE1-transduced PC-3 cells. Altogether, these data suggest that TERE1 may be significant in prostate cancer growth regulation and the down regulation or absence of TERE1 may be an important component of the phenotype of advanced disease (McGarvey et al). Recently, UBIAD1 has also been implicated to affect tumor progression in bladder cancer (Fredericks). All of the above findings showed the significance of our genetic findings and indicate that the rare variant (rs 114000606) on UBIAD1 may indicate a truly associated SNPs which infer increased risk to aggressive PCa. Men carrying risk alleles of this SNP have increased risk for developing aggressive PCa.

We then carefully calculated the study power based on our modified study design. We have >80% power to detect an OR of 2.0 (3.6) for variants with a MAF of 0.05 (0.01), at an alpha level of 1E-05 (2-sided). Therefore, we have sufficient power to identify novel rare mutations with relatively large effect based on our proposed sample size. We also considered several procedures to control for multiple test correction and SNP selection to be confirmed in additional independent samples. The Bonferroni corrected p-values are 2E-7 (0.05/200,000 variants) and 2E-6 (0.05/20,000 genes), for single variant analysis and gene-based analysis, respectively. However, not all the tests for single variants are independent due to linkage disequilibrium (LD) structure among variants. In addition, previous studies also showed that the true associations do not necessarily reach the stringent Bonferroni corrected p-value cutoffs. Therefore, to balance study power and false positives, rare variants in Aim1 that meet either of the following criteria with less stringent p-value cutoffs will be selected for replication: 1) variants reach a p-value of 1E-3 in single variant analysis; 2) variants in genes which reach a p-value of 5E-3 in gene-based analysis by SKAT. The adoption of the two-stage study design will further help to remove false positives.

In year 2, we will complete Exome Array genotyping and analysis in another 200 aggressive PCa cases and 200 indolent PCa cases, including 100 pairs of EA and 100 pairs of AA. We will combine those data with the data completed in year 1. Statistical and bioinformatics analysis will be performed in the combined dataset of 600 aggressive PCa cases and 600 indolent PCa cases. We expect to observe more variants that are significantly associated with aggressive PCa. We will follow those top significant SNPs in additional samples, as proposed in the original proposal in Aim 3.

KEY RESEARCH ACCOMPLISHMENTS

1) Completed IRB and other logistic issues
2) Performed genotyping of exome-array among 400 aggressive PCa and 400 indolent PCa in European American (EA) and AA (African American) samples
3) Performed single rare variant analysis, bioinformatics analysis, as well as gene-based analysis (SKAT) to identify rare variants that have strong effects on aggressive PCa risk

REPORTABLE OUTCOMES

1) Rare variants and genes in the genome that are significantly associated with aggressive PCa in EAs (Table 1 and Table 2)
2) Rare variants and genes in the genome that are significantly associated with aggressive PCa in AAs (Table 3 and Table 4)

CONCLUSION

1) We have made great progress in achieving the goals described in the approved statement of work

2) We have identified a list of rare variants in the genome that are associated with aggressive PCa. Those variants need to be followed in additional samples to remove false positives.

REFERENCES

Adzhubei I.A., Schmidt S., Peshkin L., et al. A method and server for predicting damaging missense mutations. Nat Methods. 7, 248-249 (2010).

Akbari MR, Trachtenberg J, Lee J, Tam S, Bristow R, Loblaw A, Narod SA, Nam RK. Association Between Germline HOXB13 G84E Mutation and Risk of Prostate Cancer. J Natl Cancer Inst. 2012 Aug 1;104(16):1260-2. Epub 2012 Jul 9.

Castro E. G.C.L., Olmos D., et al. Correlation of germ-line BRCA2 mutations with aggressive prostate cancer and outcome. ASCO Meet Abstr. 29, 1517 (2011).

Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, Wiley KE, Isaacs SD, Johng D, Wang Y, Bizon C, Yan G, Gielzak M, Partin AW, Shanmugam V, Izatt T, Sinari S, Craig DW, Zheng SL, Walsh PC, Montie JE, Xu J, Carpten JD, Isaacs WB, Cooney KA. Germline mutations in HOXB13 and prostate-cancer risk. N Engl J Med. 2012 Jan 12;366(2):141-9.

Fredericks WJ, McGarvey T, Wang H, Lal P, Puthiyaveettil R, Tomaszewski J, Sepulveda J, Labelle E, Weiss JS, Nickerson ML, Kruth HS, Brandt W, Wessjohann LA, Malkowicz SB. The bladder tumor suppressor protein TERE1 (UBIAD1) modulates cell cholesterol: implications for tumor progression. DNA Cell Biol. 2011 Nov;30(11):851-64. Epub 2011 Jul 8.

Gallagher D.J., Gaudet M.M., Pal P., et al. Germline BRCA mutations denote a clinicopathologic subset of prostate cancer. Clin Cancer Res. 16, 2115-2121 (2010).

Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, Grönberg H, Wiklund F. A Population-based Assessment of Germline HOXB13 G84E Mutation and Prostate Cancer Risk. Eur Urol. 2012 Jul 20. [Epub ahead of print]

TERE1, a novel gene affecting growth regulation in prostate carcinoma. McGarvey TW, Nguyen T, Puthiyaveettil R, Tomaszewski JE, Malkowicz SB. Prostate. 2003 Feb 1;54(2):144-55.

Thorne H., Willems A.J., Niedermayr E., et al. Decreased prostate cancer-specific survival of men with BRCA2 mutations from multiple breast cancer families. Cancer Prev Res (Phila). 4, 1002-1010 (2011).

Tryggvadottir L., Vidarsdottir L., Thorgeirsson T., et al. Prostate cancer progression and survival in BRCA2 mutation carriers. J Natl Cancer Inst. 99, 929-935 (2007).

Wu M.C., Lee S., Cai T., et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 89, 82-93 (2011).

Wan X, Yang C, Yang Q, Xue H, Tang NL, Yu W.Predictive rule inference for epistatic interaction detection in genome-wide association studies. Bioinformatics. 2010 Jan 1;26(1):30-7. Epub 2009 Oct 30.

Zhang,Y. and Liu,J.S. (2007) Bayesian inference of epistatic interactions in case-control studies. Nat. Genet., 39,1167–1173.